ENGELBERT J. DOCKNER AND MARTIN SCHEICHER*

Evaluating Volatility Forecasts and Empirical Distributions in Value at Risk Models

1. Introduction

Over the last decade financial markets have undergone significant changes. Firstly, many illiquid banking and financing instruments have been replaced by securities that are traded in national or international markets providing them with high liquidity. Secondly, the pressure for performance (i.e. the gains of individual positions or portfolios vis à vis some benchmark) has increased, thereby forcing traders and investors to take on more risk. While the risk and return trade-off is well understood in modern finance theory, too little attention has been put on risk measures and risk analysis until recently. This is partly due to the fact that the notion of risk is rather complex and that a common framework to measure risk and in particular to quantify market risk has been lacking.

Judging from the recent literature on risk analysis and management a consensus for a quantitative market risk measure seems to emerge. According to this consensus market risk is defined as the maximum potential loss of a portfolio that occurs under normal market conditions with a prespecified probability. This concept of market risk is referred to as Value at Risk (VaR) and constitutes a one-sided confidence interval on portfolio losses.[1] Although this measure has been introduced only recently it already gained enormous popularity among academics and more importantly among risk managers from the banking industry.

Risk management and VaR analyses, however, have not only received attention from academics and market participants but also from regulatory agencies like the European Community and the Basle Committee. In March 1993 the European Union published a binding Capital Adequacy Directive (CAD) in which banks and firms involved in security trading are required to hold capital according to their market risk exposure. While this directive lays out the procedures how to calculate the capital requirements when holding risky assets it does ignore the possibility of market risk

* Earlier versions of this paper have been presented at the 7th Workshop of the Austrian Working Group on Banking and Finance held at the Technical University of Vienna, October 20–21, 1995, the Workshop on Risk Management and Value at Risk for Financial Institutions held at the European Institute for Advanced Studies in Management, January 26–26, 1996 and in seminar presentations at the University of Bielefeld, the University of Magdeburg and the University of Mannheim. The authors gratefully acknowledge the useful comments of Walter Wasserfallen und Heinz Zimmermann. Department of Business and Department of Economics, University of Vienna, Brünnerstrasse 72, A-1210 Vienna.

reduction due to diversification. Therefore many EU-countries followed the suggestion of the Basle Committee (proposed in the Model Approach Paper of 1995) when implementing the directive as national law and introduced the choice for banks and security firms to either apply the CAD or an internal VaR model when calculating capital requirements. This has made VaR the standard market risk model in the entire financial industry.

VaR has gained so much popularity because it allows to represent the risk exposure of a financial institution's trading portfolio in a single number. While it is very convenient to express market risk by means of a simple statistic, there is, however, no consensus about the best implementation of this concept. In particular there are numerous approaches to calculate the VaR. They include the covariance approach, historical and Monte Carlo simulation. The reason why there is no unique approach is rooted in the definition of VaR. The analytical definition of the VaR of a trading portfolio reveals that it depends on (i) the specification of the distribution of the portfolio returns, (ii) the valuation model that is used to map nonlinear positions into their underlying securities, and (iii) a volatility estimate for portfolio returns. Since the theoretical and empirical finance literature offers a variety of alternative models and concepts to specify these individual parts of a VaR model, there is no unique VaR method. While this might be attractive from an academic point of view, it may cause uncertainty (and even confusion) among practitioners who generally like to apply a single method to as many different portfolios as possible.

One possible way to get a better understanding of the different methods is to empirically evaluate their performance as risk measurement systems. In fact, a lot of current research is devoted to this task. Existing papers evaluate the performance of historical and Monte Carlo simulation (see PRITSKER (1996), BUTLER and SCHACHTER (1996)), evaluate the distribution assumptions for VaR calculations (see FALLON (1996)), and deal with the verification of the accuracy of VaR models (see KUPIEC (1996)).

In this paper we take up the issue of empirically evaluating the performance of VaR models and focus on two of the above mentioned components of such a model. We evaluate the forecasting performance of different volatility models that are required for calculating confidence intervals and we study the significance of the returns distribution assumption for the accuracy of VaR. We analyze the forecasting performance of three commonly used volatility models: (i) constant volatility, (ii) exponentially weighted moving average (EWMA), and (iii) GARCH volatility. As for the distribution assumptions we compare three alternative scenarios. One in which the VaR is calculated on the basis of a normal distribution for the portfolio returns, one in which we make use of Student's t distribution and finally a GARCH model where the conditional returns are assumed to follow Student's t.

Our main results are that while the choice of a specific volatility model has on average no significant influence on the VaR estimate (the forecasting performance of the three models are quite similar), the choice of the returns distribution does have a significant consequence for the VaR estimate. The latter result is not surprising, since different distributions have different kurtosis and hence predict the tail events (which are of interest in risk management) differently.

The paper is organized as follows. In the next section we briefly discuss the concept of VaR and the different methods of implementing it. In Section 3 we discuss alternative volatility models for quantifying market risk. In Section 4 we describe our data set and evaluate the forecasting performance of the alternative volatility models. Section 5 analyses the impact of the distribution assumption on the VaR estimate and evaluates different specifications empirically and finally Section 6 concludes the paper.

2. Value at Risk

According to the standard definition Value at Risk is the maximum amount one can expect to lose on a given position during a given period (the potential close-out period) with a predefined probability (cf. JORION (1997) and RISKMETRICS™ (1994)). This definition is based on a statistical approach to quantify market risk and identifies the VaR of a given portfolio with a one sided confidence interval on its potential losses.
Let S_t be a random state variable that measures a stock price in period t for example. $\Delta S = S_t - S_{t-\Delta t}$ is the change of the state variable over the period Δt. Let $V(S_t,t)$ be the value of a portfolio at time t. If this portfolio includes derivative securities with an underlying asset S its value V does depend on the current price of the underlying. Over a period of length Δt the value of the portfolio changes by $\Delta V(\Delta S, \Delta t)$, where ΔS is the change in the price in the underlying asset and Δt is the holding period or forecast horizon. Based on the change ΔV we can define the value at risk of the portfolio. It is given by (cf. FALLON (1996)):

$$\mathrm{Prob}[\Delta V(\Delta S, \Delta t) > -\mathrm{VaR}] = 1 - \alpha \qquad (1)$$

where α is the confidence level.
From this definition it is clear that the calculation of the VaR of a portfolio depends on the following elements: (i) the distribution of the returns (changes) of the portfolio V, (ii) the pricing model that associates a value for the nonlinear position in the portfolio with the underlying instruments. Finally, (iii) the VaR does also depend on the volatility of the asset S. In order to see these dependencies let us first consider the case of a linear portfolio, i.e. a portfolio that consists of a position in the asset S only where ΔS is normally distributed with zero mean and constant variance σ^2, $\Delta S \sim N(0, \sigma^2)$. Then we get

$$\Delta V = \delta \, \Delta S \qquad (2)$$

and the value at risk for a confidence level of α is given by

$$\begin{aligned}\mathrm{Prob}[\Delta V(\Delta S, \Delta t) > -\mathrm{VaR}] &= \mathrm{Prob}[\delta \Delta S > -\mathrm{VaR}] \\ &= \mathrm{Prob}[\Delta S/\sigma > -\mathrm{VaR}/\delta\sigma] = 1 - \alpha \qquad (3)\end{aligned}$$

Since ΔS is normally distributed the change in the portfolio value is also normally distributed and the 100(α)th percentile, $Z(\alpha)$, of the standard normal distribution is equal to

$$Z(\alpha) = -\mathrm{VaR}/\delta\sigma. \qquad (4)$$

This results in a VaR given by

$$\mathrm{VaR} = -Z(\alpha)\, \delta\, \sigma \qquad (5)$$

From equation (5) it becomes obvious that in order to forecast the value at risk for the next day or the holding period Δt, a volatility estimate for the asset S is required. This also holds true in the case of nonlinear positions like derivative securities. Nonlinear positions require a pricing model to relate the price of the underlying to the price of the derivative security. Using the same notation as above, $V(S_t)$ is the value of a stock option, for example, with the current stock price given by S_t. Making use of linear approximations the change in the value of the option ΔV can be related to the change in the stock price identically as in equation (2) where δ is the option delta. Hence it becomes obvious that the same approach as in the case with a stock portfolio can be applied under the restriction, however, that the nonlinearity of the option position is linearized with the corresponding approximation error. But again it becomes obvious that the forecast of the stock volatility is necessary together with an assumption on the return distribution and a valuation model (with the corresponding δ) to calculate the VaR.
Although market risk can elegantly be expressed in a single statistic as in equation (5), there exist numerous ways to forecast volatility or to specify

the returns distribution. Only empirical analysis can shed light on the problem of finding the most appropriate model. In this paper we concentrate on the issues of volatility forecasts (σ) and distribution estimation ($Z(\alpha)$). In particular we want to empirically evaluate the influence of the volatility forecast and the distribution specification on the VaR. For that matter we briefly introduce different volatility concepts in the next section.

3. Choosing A Model for Volatility

The literature on the econometric modeling of financial time series does not contain a standard and well accepted definition of volatility. For recent surveys illustrating the variety of models available see PALM (1996) or DIEBOLD and LOPEZ (1995). In this paper we do not want to give a literature survey on volatility models. We rather concentrate on those specifications that are employed in this study. Hence our selection of volatility specifications is driven by their forecasting abilities.[2]

a) Naive model: The benchmark for measuring the forecasting performance of any volatility model is the updated sample variance, defined below. Under the assumption of normality the distributions of stock returns can fully be described by the mean and the variance. While this is appealing, the naive model neglects the above mentioned stylized facts like volatility clustering, and fat tails. In algebraic form the naive model can be written as

$$\sigma_t^2 = \frac{1}{k-1}\sum_{i=1}^{k}(r_{t-i} - \mu)^2 \qquad (6)$$

where r_t is the compounded rate of return ($r_t = \ln X_t - \ln X_{t-1}$), μ is the mean return and k is the sample length. For high frequency time series (i.e. daily stock returns) the mean is usually close to zero hence can be neglected.

b) Exponentially Weighted Moving Average: The main disadvantage of the naive model is that it gives equal weight to all observations in the sample, thus neglecting the stronger impact of recent innovations. This is the reason why the naive model is not capable of mimicking volatility clustering present in stock returns. This has led to the introduction of alternative moving average models for computing stock return volatilities (see for example TAYLOR (1986)). Comparing several specifications, Taylor found that the Exponentially Weighted Moving Average (EWMA) did best with respect to empirical performance. The idea behind the EWMA is the following. The volatility of the current period t is calculated as an MA(∞) process of weighted squared deviations from the mean where the weights decay exponentially, i.e.,

$$\begin{aligned}\sigma_t^2 &= (1-\lambda)\sum_{i=0}^{\infty}\lambda^i(r_{t-i-1} - \mu)^2 \\ &= \lambda\sigma_{t-1}^2 + (1-\lambda)(r_{t-1} - \mu)^2\end{aligned} \qquad (7)$$

where λ is the weight or decay factor.

The definition of the EWMA model makes apparent that it is a generalization of the standard variance estimator with decaying weights. While the definition of the EWMA is rather straightforward its use depends on the estimation of the weight λ. It can be chosen by minimizing an appropriate error measure. What is also seen from specification (7) is that an initial value for the series is needed in empirical estimation. Usually the sample variance is chosen. The volatility estimates of RISKMETRICS™ are based on an EWMA-model. But in the empirical specification the decay parameter λ is not estimated using some objective function but it is arbitrarily set equal to .94. All the volatility forecasts are based on this specification. When we evaluate the forecasting abilities of the different specifications we will also use this value.

c) GARCH models: The most successful model for describing nonlinear dynamics and non-normality of stock returns is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, introduced by ENGLE (1982). It is a model that particularly builds on time-varying second order moments. Applications of GARCH in finance are surveyed in BOLLERSLEV et al. (1992). While the class of GARCH models is rather flexible and admits a large variety of different specifications we concentrate on GARCH(1,1) that is defined as follows

$$r_t = \mu + u_t \tag{8}$$

with $$u_t = \varepsilon_t \sqrt{h_t} \tag{9}$$

with $\varepsilon_t \sim N(0,1)$ and

$$h_t = a_0 + a_1 u_{t-1}^2 + a_2 h_{t-1} \tag{10}$$

This specification implies that the conditional variance follows an ARMA-type process where stationarity is satisfied when the sum of a_1 and a_2 is less than 1. If this condition is violated there is a unit root in the variance process and the corresponding model is referred to as an integrated GARCH (IGARCH) model. The introduction of conditional time-dependent second moments is capable of generating all stylized facts of stock returns mentioned in section 1. In particular we obtain volatility clustering as well as leptokurtotic unconditional returns. This holds true even when a Gaussian distribution is specified for the standardized residuals ε_t.

Comparing the GARCH model for quantifying market volatilities with the EWMA specification introduced above we observe that the two are closely related. The major differences stem from the way the model parameters are estimated and from the their stationarity assumptions. As for the parameter estimation GARCH models do make use of statistical techniques like maximum likelihood estimation whereas the EWMA as used in RiskMetrics™ arbitrarily sets the weight equal to .94. Moreover the GARCH model includes a constant term in its variance specification and does not assume nonstationarity from the outset. But apart from these differences the two models are quite similar and hence we can expect them to deliver similar volatility forecasts. Since EWMA assumes a nonstationary variance process we can expect that this model does well when actual data are close to being nonstationary. Nevertheless we argue that the GARCH framework is the more flexible model since it offers many alternative specifications and its parameters are estimated based on observed data. For a survey on different GARCH specifications see PALM (1996). The most popular extensions to the standard GARCH model as introduced here is the exponential GARCH (EGARCH), proposed by NELSON (1991) and the GLOSTEN, JAGANNATHAN, RUNKLE (1993) model. These modifications do offer improved fits relative to the standard model but they can only be applied to stock returns. Since a methodology for estimating market risk requires some flexibility as far as the application to varies securities is concerned we do not apply EGARCH models in this paper.

d) *Implied Volatility:* The models outlined so far all use time-series methods to forecast market volatility. An alternative approach is to make use of option prices and calculate what is referred to as implied volatility. The idea behind this approach is the following. For many securities traded in the market derivative products like options do exist that are priced on a daily basis. Since option prices are determined by the volatility of the underlying asset these prices together with an option price model can be used to calculate implied volatilities which can serve as volatility estimates. While this approach is appealing it suffers from several major drawbacks. Firstly, options do not exist for all the securities traded in a market and hence the concept of implied volatility is not universally applicable. Secondly, the implied volatilities do depend on the option pricing model. These two arguments lead us not to pursue the concept of implied volatilities any further. A full discussion of this issue is given in LAMOUREUX and LASTRAPES (1993) or BATES (1996).

4. Volatility Estimates and Forecasts

4.1 Statistical Properties of Sample Data

Our sample comprises 1500 daily compound returns of three stock market indices and an interest rate future: DAX, CAC, FTSE and the German Government Bond Future, GeGB. We choose the daily frequency as it is most useful for volatility models. High frequency (= intraday) samples contain too much noise and cause a heavy computational burden and monthly data result in too few observations. Our sample was partioned into the first 1,000 observations in-sample for modeling and the remaining 500 observations were used for out-of-sample forecasts. The sample period is given by 89/1/6 to 95/7/17.

Figure 1: Normalized Estimated Unconditional Distribution of GEGB

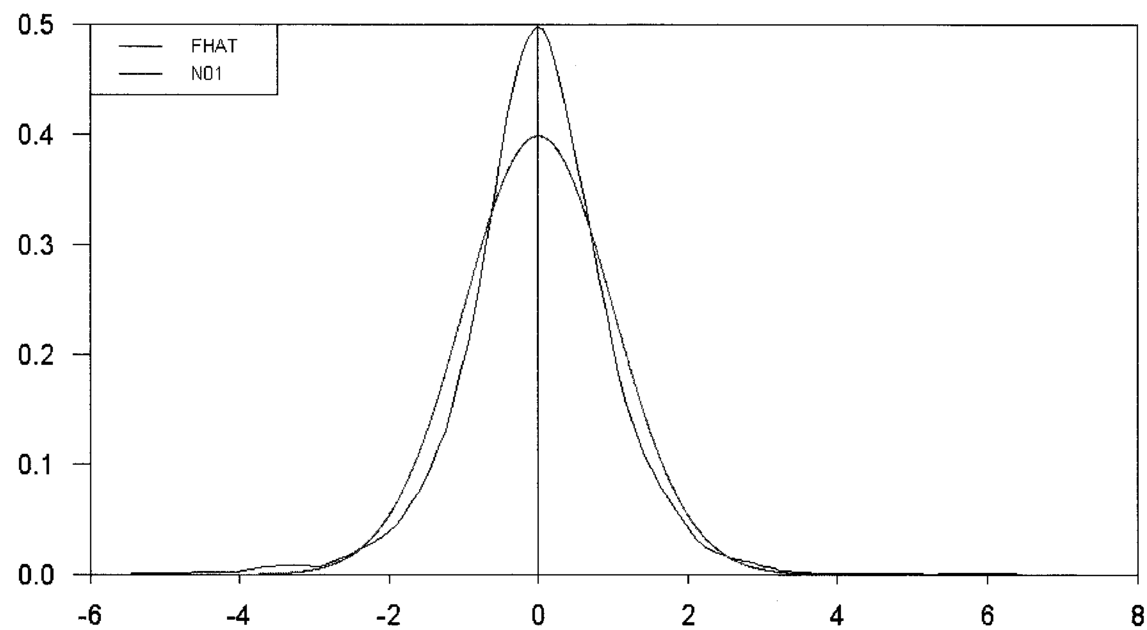


Table 1: Descriptive Statistics

	DAX	CAC	FTSE	GeGB
MEAN * 1000	0.2779	0.1083	0.3811	0.01099
VARIANCE * 1000	0.131	0.123	8.231	0.01284
SKEWNESS	–1.1211*	–0.2331*	0.1714*	–0.0632 (0.28)
KURTOSIS	18.4161*	3.6722*	2.3623*	3.9065*
Q(24) of r_t	20.09 (0.69)	22.00 (0.57)	35.02 (0.06)	48.96*
Q(24) of r_t^2	63.88*	176.68*	133.00*	722.78*

This table presents the first four moments; Q(24) is the Ljung Box Q statistics for joint significance of the first 24 autocorrelations, * indicates significance level of 99% and above.

Table 1 summarizes some descriptive statistics for the four returns series. These statistics show familiar behavior. There is significant nonnormality, in particular excess kurtosis. Heteroscedasticity in the form of strong volatility clustering is responsible for the fat tails. Moreover the Q-statistics indicate that there is autocorrelation and nonlinearity in the data. To get another perspective on the distribution of the sample series, Figure 1 contains a nonparametric estimate of the normalized unconditional distribution of GEGB returns[3] (FHAT) and the N(0,1) distribution for comparative purposes. The nonparametric estimation is done with an Epanechnikov kernel. Again nonnormality is evident.

4.2 In-Sample Estimation of Volatility Models

We start our econometric evaluation by estimating a GARCH (1,1) model for each of the four time series with the sample from 89/1/6 to 95/7/17. The model is estimated by optimizing the Maximum Likelihood Function with a BHHH algorithm. Results are summarized in Table 2.

Table 2: GARCH(1.1) results 89/1/6 to 95/7/17

$$r_t \mid I_{t-1} \sim N(\mu, h_t)$$

$$h_t = a_0 + a_1 u_{t-1}^2 + a_2 h_{t-1}$$

Variable	DAX	CAC	FTSE	GeGB
μ * 1000	0.4821 (0.06)	0.1977 (0.44)	0.03951 (0.04)	0.01507 (0.82)
a_0 * 1000	0.01015*	0.0939*	0.003455*	0.000111 (0.25)
a_1	0.1378*	0.0928*	0.0692*	0.0797*
a_2	0.7991*	0.8319*	0.8809*	0.9145*
LRT IGARCH (1)	24.96*	19.26*	11.61*	1.10 (0.29)
Degrees of freedom (2)	5.4159*	7.2974*	10.8975*	7.1731*

* significance level of coeffiecient is 99% and above.
(1) Likelihood Ratio Test for Integrated GARCH. i.e.: H_0 is $a_1 + a_2 = 1$
(2) Degrees of freedom estimated for GARCH with conditional Student's t distribution instead of Normal

They show significant GARCH effects. To investigate the evidence of a unit root in the second moment, we computed the Likelihood Ratio Tests (= LRT) for the restriction of an integrated GARCH model ($a_1 + a_2 = 1$). All four returns' series reject the IGARCH specification. As a second volatility model we make use of the EWMA as specified in RiskMetrics™. This implies that we fix the decay factor λ and set it equal to 0.94. Because the coefficients for the EWMA model are fixed a priori we can not directly compare this model to a GARCH model by computing specification tests. Still we can note that the estimated GARCH coefficients differ from the values 0.94 and 0.06, which are used in RiskMetrics™. Of particular importance is also that the constants a_0 in the conditional variance equation of the GARCH model are significant (cf. Table 2). Assuming that the true volatility model is of the GARCH-type, this difference will cause biased forecasts for the EWMA model.

4.3 Out of Sample Forecasting Performance

Since our objective is to evaluate the forecasting performance of alternative volatility models and their consequences for VaR estimates we make use of the following methodology. We start with a sample of 1000 observations for estimating the sample variance and the coefficients for the GARCH(1,1) specification for each of the four return series. Based on these estimated structural equations, forecasts are computed for the next k = 10 periods by making use of the following volatility equations:

GARCH(1,1)

$$(k = 1: \quad h_{t+1} = a_0 + a_1 u_t^2 + a_2 h_t \tag{11}$$

$$k > 1: \quad h_{t+k} = a_0 + (a_1 + a_2) h_{t+k-1}$$

EWMA:

$$k = 1: \quad \hat{h}_{t+1} = 0.06 u_t^2 + 0.94 h_t \tag{12}$$

$$k > 1: \quad \hat{h}_{t+k+1} = \hat{h}_{t+k}$$

Naive Model:

$$\hat{h}_{t,t+k} = \frac{1}{999} \sum_{i=1}^{1000} (r_{t-i} - \mu)^2 \tag{13}$$

Forecasting is continued by discarding 10 observations at the beginning and adding 10 at the end. Again GARCH and the sample variance are reestimated and new forecasts are computed. This method of rolling forecasts leaves a constant sample length of 1000 observations. An alternative to the rolling sample is the updated sample method, which grows over time. We believe that rolling forecasts are more useful as the influence of innovations from the past is diminished. In ENGLE et al. (1993) a GARCH model with 1000 observations gives the best performance when comparing models in a simulated options market. Overall we obtain 10 times 50 = 500 forecasts. We compared the behavior of estimated GARCH coefficients across the rolling subperiods. Their values are quite close to those reported in Table 2 with some differences after large price changes.

To evaluate the forecasting performance of the alternative volatility models we require a benchmark which in our case is assumed to be given by the squared daily returns. Although this assumption is critical, there is no alternative as the literature on forecasting volatility documents, e.g. PAGAN and SCHWERT (1990), SCHMITT (1994), WEST and CHO (1994), JORION (1995) or FIGLEWSKI (1997).

The evaluation of the forecasting performance of the three different models is based on two different statistical tests: the Root Mean Squared Error (RMSE) and a linear regression approach introduced by PAGAN and SCHWERT (1990).

Table 3a: Mean Errors of 10-day Volatility Forecasts[1)]

	DAX	CAC	FTSE	GeGB
Const.Vola.	3.8135e–5	2.3137e–5	1.7056e–5	–4.9143e–6
EWMA	–9.0271e–5	–3.7089e–6	–7.8438e–7	1.1214e–7
GARCH	2.8879e–5	1.3804e–5	1.0522e–5	–2.4229e–6
IGARCH	6.8627e–5	5.6447e–5	2.0621e–5	1.9932e–6

[1)] 1000 observations are used for model estimation. The mean error is defined as $\frac{1}{500}\sum_{t=1}^{500}\left(e_t^2 - h_t\right)$
Const Vola. is the model with constant volatility, the other models are described in the text;

Table 3b: Root Mean Squared Errors of 10-day Volatility Forecasts[1)]

	DAX	CAC	FTSE	GeGB
Const.Vola.	1.6243e–4	1.4760e–4	7.8807e–5	2.8610e–5
EWMA	1.5472e–4	1.4566e–4	7.4964e–5	2.6362e–5
GARCH	1.6145e–4	1.4502e–4	7.5592e–5	2.6432e–5
IGARCH	1.7862e–4	1.5833e–4	8.0381e–5	2.6878e–5

[1)] 1000 observations are used for model estimation. The root mean squared error is $\left[\frac{1}{500}\sum_{t=1}^{500}\left(e_t^2 - h_t\right)^2\right]^{0.5}$
Const Vola. is the model with constant volatility, the other models are described in the text;

As the first statistical measure to evaluate the forecasting performance of the volatility models we use the RMSE. It is defined as:

$$\mathrm{RMSE} = \left[\frac{1}{500}\sum_{t=1}^{500}\left(e_t^2 - h_t\right)^2\right]^{0.5} \tag{14}$$

where e_t^2 refers to the realized volatility, i.e. the benchmark, and h_t to the forecast. The definition of the RMSE implies that it measures the spread of the estimates around the true value and that it is the sum of the squared bias and the variance of the estimator. On the contrary the Mean Error is a measure of this bias only.

Table 3 reports the ME and RMSE for the different volatility models with a rolling window of 10 days. It also includes the RMSE of the IGARCH specification for all four time series. According to the RMSE measure the EWMA model does best in three of the four cases (DAX, FTSE and GeGB) while the GARCH specification outperforms the other models for the CAC. It must be noted, however, that the results are very close and that there is no substantial difference between all the three different models. In order to get additional insights into the forecasting performance of the volatility models we have also calculated the RMSE for the case with a rolling window of one trading day. The corresponding results are given in Table 4.

They also show similar behavior but now the EWMA ranks best for all four time series. Here again the differences for the alternative models are small.

We have pointed out that the RMSE is the sum of the squared bias and the variance of the estimator. It is, however, not possible to get information on the source of the bias of the estimator. This problem is discussed in FAIR and SHILLER (1990) in the context of forecasting changes in GDP. As an alternative method we apply the tests proposed by PAGAN and SCHWERT (1990) for volatility models. Instead of looking for the lowest RMSE, they regress realized volatility on a constant and the forecasted volatility by means of a linear model

$$e_t^2 = c_0 + c_1 h_t + \varepsilon_t \tag{15}$$

According to this test if $E(e_t^2) = h_t$ (no bias), c_0 should equal 0, c_1 should equal 1, and the residuals should show no significant autocorrelation. Based on this linear regression we can consider the t-statistic of the coefficients as a measure for the bias, and the R^2's as a measure of the variance of a forecast. As already mentioned the bias and the variance are also the components of the RMSE but in that case they are given in aggregate form whereas with the linear regression they can be separated out. Equation (15) is estimated by ordinary least squares, however we make an adjustment for the variance-covariance matrix to get a consistent estimator for standard errors. Here we choose the method proposed by WHITE (1980): The standard errors are made robust

Table 4a: Mean Errors of 1-day Volatility Forecasts[1)]

	DAX	CAC	FTSE	GeGB
Const. Vola.	3.8135e–5	2.3250e–5	1.7050–5	–4.9143e–6
EWMA	1.047e–6	–4.6881e–6	–1.1260e–6	1.0817e–7
GARCH	1.8479e–5	9.2554e–6	7.4886e–6	–1.9549e–6

[1)] 1000 observations are used for model estimation. The mean error is defined as $\frac{1}{500}\sum_{t=1}^{500}\left(e_t^2 - h_t\right)$

Const Vola. is the model with constant volatility, the other models are described in the text;

Table 4b: Root Mean Squared Errors of 1-day Volatility Forecasts[1)]

	DAX	CAC	FTSE	GeGB
Const. Vola.	1.6243e–4	1.4765e–4	7.8807e–5	2.8610e–5
EWMA	1.5526e–4	1.4649e–4	7.5294e–5	2.6335e–5
GARCH	1.5787e–4	1.4660e–4	7.5965e–5	2.6496e–5

[1)] 1000 observations are used for model estimation.

The root mean squared error is defined as $\left[\frac{1}{500}\sum_{t=1}^{500}\left(e_t^2 - h_t\right)^2\right]^{0.5}$

Const Vola. is the model with constant volatility, the other models are described in the text;

Table 5a: Linear Regressions for 1- day GARCH Forecasts

$$e_t^2 = c_0 + c_1 h_t + \varepsilon_t$$

	DAX	CAC	FTSE	GeGB
Constant*10⁴	0.6272 (1.69)	0.9403 (2.48)	0.0690 (0.47)	0.0138 (0.91)
Beta	0.2576 [0.31]	0.1249 [0.30]	0.7648 [0.21]	1.0416 [0.068]
R^2	0.0026	0.0003	0.021	0.0975

* indicates a significance level of 99% and above; (t-statistic) [standard error]

Table 5b: Linear Regressions for 1- day EWMA Forecasts

$$e_t^2 = c_0 + c_1 h_t + \varepsilon_t$$

	DAX	CAC	FTSE	GeGB
Constant*10⁴	0.4475 (2.28)	0.8421 (3.74)	0.1821 (2.81)	0.0228 (2.22)
Beta	0.5019 [0.18]	0.236 [0.20]	0.675 [0.11]	0.8473 [0.04]
R^2	0.0151	0.0018	0 .0358	0.1069

* indicates a significance level of 99% and above; (t-statistic) [standard error]

against heteroscedasticity and against autocorrelation up to lag 36. The advantage of this method is that we can gain information on the trade-off between the variance and the bias of a forecast and the rankings according to the two methods can be compared. The results based on the linear regression test are given in Table 5. Since according to the RMSE the best forecasting performance was achieved either by the GARCH or the EWMA we restrict the regression test only on those models. First, let us look at the performance ranking according to the R^2.
Here we find that the EWMA is superior to GARCH for all four returns series. Second, we focus on the t-statistics. Here GARCH is superior to EWMA in all cases.
Based on these results we can draw the following conclusions. The forecasting performance of the EWMA is superior to the GARCH specification in all four cases but the GARCH(1,1) model has a lower bias. Again we want to point out, however, that the differences in the two models are quantitatively very small. In order to analyze the implication of this result for the accuracy of a VaR estimate we proceed as follows. For each of the for returns series we start out with a hypothetical portfolio with the value of 1.000 on January 2, 1993 and forecast the VaR for this position on a daily basis by making use of the different volatility forecasts together with the assumption of normally distributed returns and a confidence level of 95%. This results in VaR forecasts given by

$$VaR_t = 1.000 * 1.65 * \sigma_t \quad (16)$$

where σ_t is the volatility forecast for the next day. The plots of the VaR with GARCH and EWMA are shown in Figure 2 and 3 for GeGB.[4] It

Figure 2: VaR of GARCH & Change of Portfolio

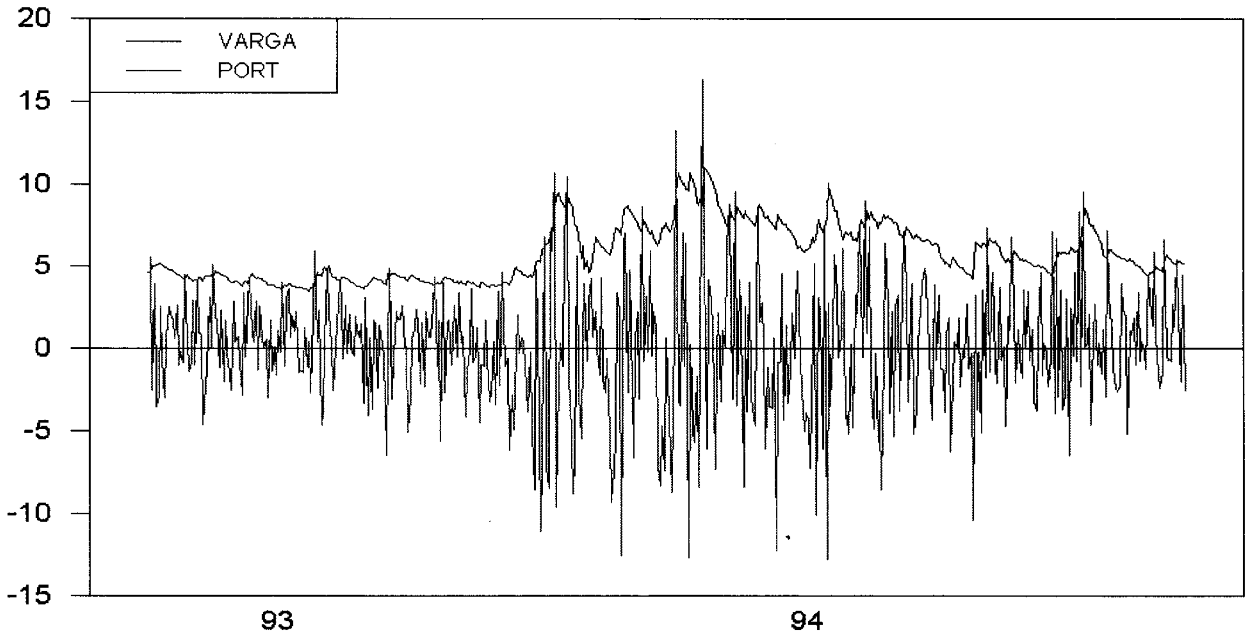


Figure 3: VaR of EWMA & Change of Portfolio

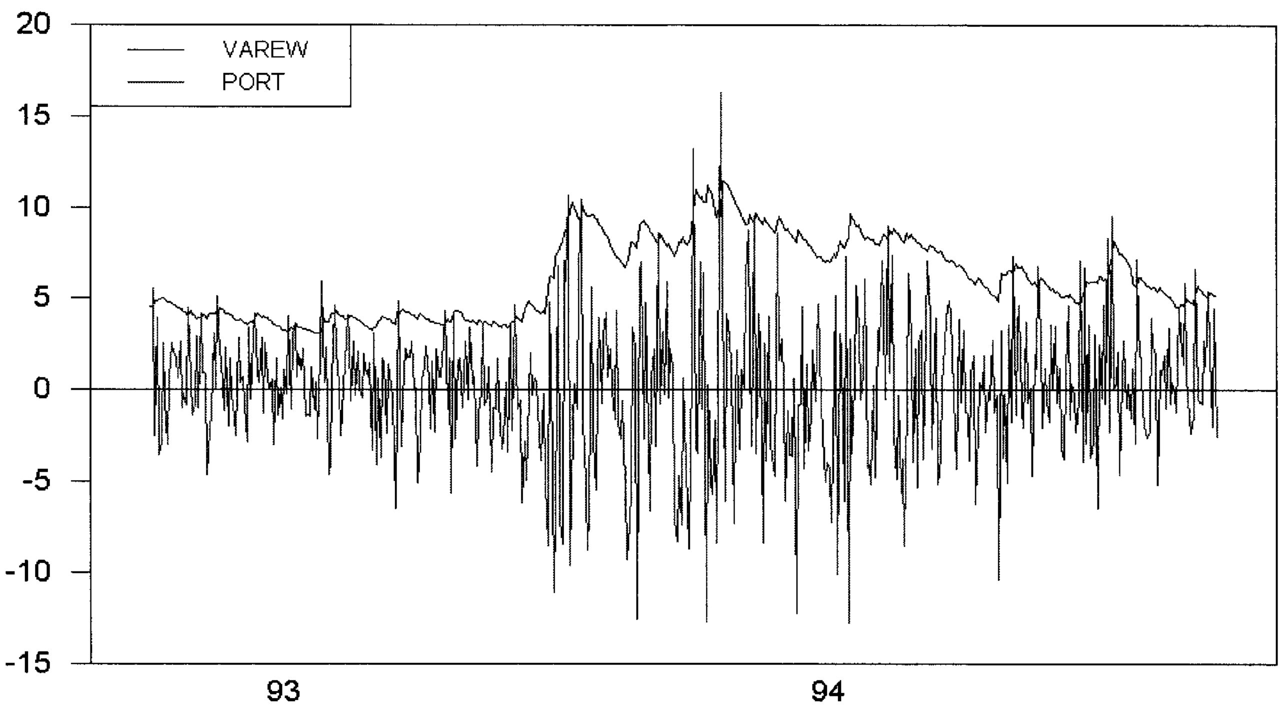


Figure 4: VaR-t of GARCH & Change of Portfolio

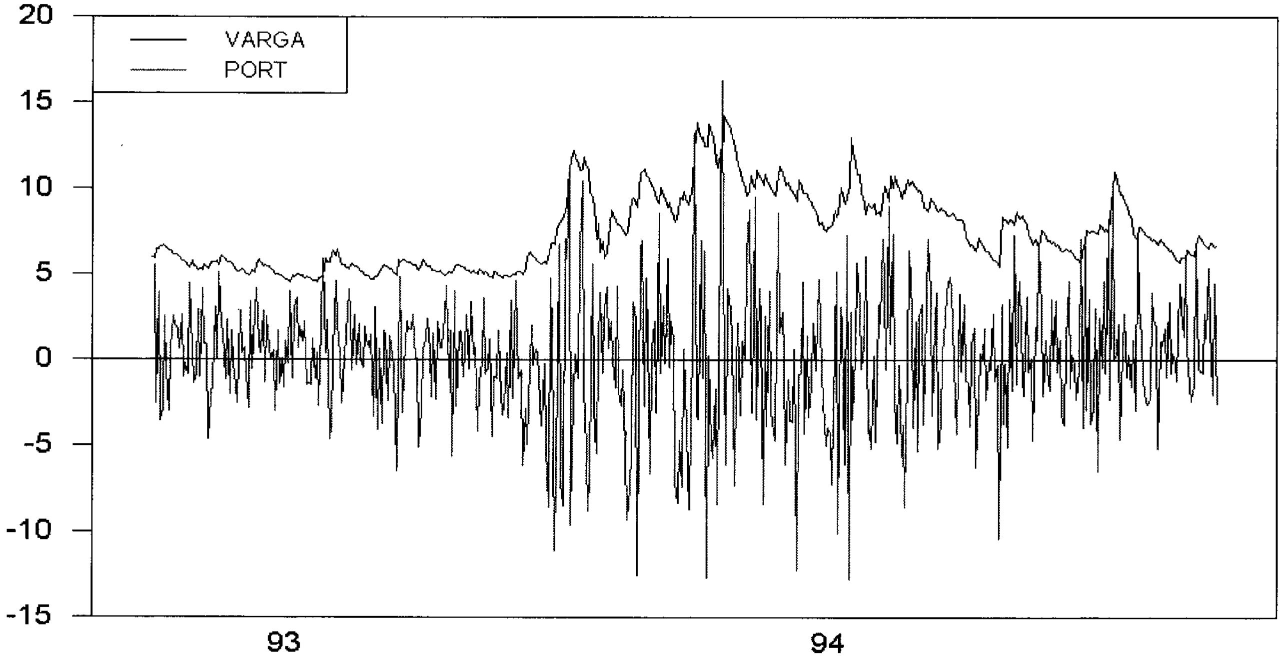
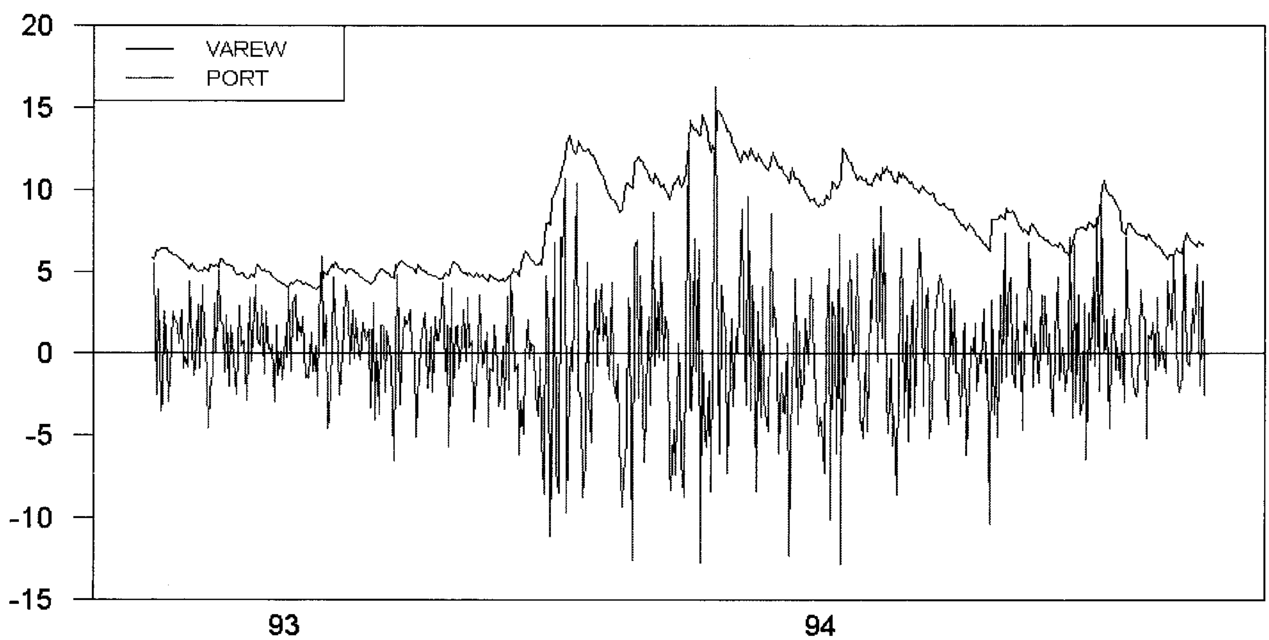


Figure 5: VaR-t of EWMA & Change of Portfolio



strongly supports our findings for the performance of the volatility estimates and lead us to one of the main conclusions of the paper: The choice of the volatility model does not have a strong impact on the performance of a value at risk based risk management system.

Moreover in Table 6 we present the number of times the actual loss of the portfolio was larger than the forecasted one. We find that when the VaR is calculated on the basis of a normal distribution and EWMA volatility forecasts the risk is systematically underestimated while in case of GARCH volatilities the number of outliers corresponds roughly with the significance level of 95%. This findings support the fact that taking into account leptokurtic distrubutions is crucial for the estimation of the VaR. While GARCH models are capable of capturing some leptokurtosis present in the data simple EWMA models do not, which explains the difference in the results.

5. Distribution Estimation and VaR

It is evident from the test statistics given above that daily return series are NOT normally distributed. This is already the common finding since the 1960´s, when Eugene Fama and Benoit Mandelbrot started the literature on the empirical characteristics of financial time series. More recent examples are among many others HSIEH (1988) for Pound Sterling, Canadian $, DM, Swiss Franc and Japanese Yen against US $, KIM AND KON (1994) for 30 US stocks and 3 stock indices and all these papers use samples at a daily frequency. Since the Gaussian allocates too little probability mass in the tails the chance of a crash is underestimated. But when estimating market risk, it is exactly the extreme values which are of particular importance.

An answer to the question of finding the appropriate distribution for asset returns can be found in PEIRO (1994). He compares several alternative specifications on daily returns from 6 stock market indices: Using Pearson´s goodness-of-fit statistic and Likelihood ratio tests for nested distributions he evaluates the Student´s t, Paretian, mixture-of-normals, Logistic and Exponent power with the Normal as a benchmark. His result is that the Student´s t offers the best fit. This is intuitive, because by means of an estimate for the degrees of freedom it can accomodate the observed excess kurtosis. For degrees of freedom which are higher than 10, the Student´s t is close to the Normal. Table 7 gives the estimated degrees of freedom for the Student´s t distribution for our sample series. We again reject the Gaussian distribution, as all estimates are below 5. The improvement of the Student´s t respective to the Gaussian can be quantified using posterior probability criteria. We computed the Schwarz Information Criterion for both specifications. They are given in Table 8. All show that as a static or time-invariant model for returns, the t distribution offers a significant improvement over the Normal. The same is true for GARCH as a time-dependent model.

Table 6: Number of Outliers of VaR models

	DAX	CAC	FTSE	GeGB
EWMA	45	32	40	24
GARCH	36	24	15	11
EWMA–t	16	12	24	6
GARCH–t	13	9	12	7

Table 7: Student's t estimation results 89/1/6 to 95/717

	DAX	CAC	FTSE	GeGB
Mean*1000	0.3732	0.1872	0.3570	0.0486
Variance*1000	0.1222	0.1228	0.06687	0.00133
Degrees of F.	3.9682	5.3512	8.3204	3.9979

* indicates significance level of 99% and above.

Table 8: Schwarz Information Criteria

SIC = L – 0.5*log(N)*k

	DAX	CAC	FTSE	GeGB
Normal	5186.1526	5234.2013	5747.2589	7162.6696
Student's t	5384.3615	5300.8409	5779.0098	7264.7793
GARCH	5283.3979	5292.8587	5780.6112	7359.0807

We can now go back to the definition of the market or Value at Risk. Under Normal distribution and a confidence level of 95 % this is[5]:

$$VaR_m = MV\ 1.65\ \sigma_m \quad (17)$$

Making use of the results from fitting the Student´s to the series in our sample we get the following results:

$$VaR_{DAX} = MV\ 2.13\ \sigma_{DAX} \quad (18)$$
$$VaR_{CAC} = MV\ 2.01\ \sigma_{CAC} \quad (19)$$
$$VaR_{FTSE} = MV\ 1.86\ \sigma_{FTSE} \quad (20)$$
$$VaR_{GeGB} = MV\ 2.13\ \sigma_{GeGB} \quad (21)$$

The differences in the results are high, as the values from the Normal and the t distributions only converge for very high degrees of freedom. The advantage of this method is that it accounts for the asset-specific leptokurtosis. This is a considerable improvement on the rigid assumption of Normality.

Based on the estimation of empirical distributions we report in Table 6 the number of times the actual loss of the hypothetical portfolio is larger than the estimated one. Here we find that risk is either correctly quantified or even overestimated.

6. Conclusions

The econometric evaluation of several market risk models proposed in this paper allows the following conclusions. In the Technical Document the EWMA is chosen to forecast standard deviations and the Gaussian distribution is assumed. Based on the above results we propose the GARCH class of models and for the unconditional distribution we select the Student´s t. GARCH nests the EWMA and is so the more flexible approach. For stationary series GARCH has a better performance. It is not convincing to make the assumption of nonstationarity for all the assets in a portfolio. This can only be handled in a

GARCH framework. It has to be pointed out however that for nonstationary series EWMA gives a good performance. The added advantage of the EWMA is that it is easy to apply. Thus for volatility forecasting the RiskMetrics™ approach is on the whole a satisfying method. One caveat in this conclusion is the choice of selection criteria. During the estimation we have found a strong dependence of the results on the criteria used. In particular the assumption that realized volatility is given by the squared series is not supported by economic theory, which makes it questionable. As we have already pointed out the results for RMSE are sometimes very close and there is clear solution.

However we have to criticize the assumption of a normal distribution. The t distribution gives a better fit to the return series and it allows for an asset-specific measure for fat tails. This characteristic is especially important in the context of the discussion about the panic factor. The proposal of the European Union for the Capital Adequacy Directive was to allow banks to use their individual models. However the results should be scaled by a factor of three to make them more robust: This was dubbed the "panic factor". With the t-distribution there is no need for this restrictive measure. Here the degrees of freedom measure the probability mass in the tails of the density of a series. As we have seen above the resulting VaR estimates differ strongly.

Footnotes

[1] For a detailed definition of the VaR measure see JORION (1997) or the Technical Document of RiskMetrics™.

[2] See also BRAILSFORD and FAFF (1996) or FIGLEWSKI (1997).

[3] Results for other series are similar.

[4] Results for other series are similar.

[5] RiskMetrics™ also assumes that the cutoff point is set for the 5 % tail area. In their comments on RiskMetrics™ LAWRENCE and ROBINSON (1995) discuss the choice of the cutoff point. They are in favor of a tighter value, namely 1%. However as the choice of 5% is common in statistical inference, we have decided to keep that value.

References

AKGIRAY, V. (1989): "Conditional Heteroscedasticity", Journal of Business 62(1), pp. 55–80.

BATES, D. (1996): "Testing Option Pricing Models" in: Maddala, C. S. and C. R. Rao (eds): Handbook of Statistics 14, pp. 567–612.

BOLLERSLEV, T. (1986): "GARCH", Journal of Econometrics 31, pp. 307–327.

BOLLERSLEV, T. , R. CHOU and K. KRONER (1992): "ARCH Modeling in Finance", Journal of Econometrics 52, pp. 5–59.

BOLLERSLEV, T., R. ENGLE and D. NELSON (1993): "ARCH Models", Discussion Paper 93–49, UCSD.

ENGLE, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation", Econometrica 50, pp. 987–1008.

ENGLE, R., C. HONG, A. KANE and J. NOH (1993): "Arbitrage Valuation of Variance Forecasts with Simulated Options", in: Chance, D. and R. Trippi (eds.): Advances in Futures and Options Research, JAI Press.

BRAILSFORD, T. and R. FLAFF (1996): "An evaluation of volatility forecasting techniques" Journal of Banking & Finance 20, pp. 419–438.

DIEBOLD, F. and J. LOPEZ (1995): "Volatility Models", Federal Reserve Bank of New York Research Paper #9522.

FAIR, R. and R. SHILLER (1990): "Comparing Information in Forecasts from Econometric Models", American Economic Review, pp. 375–389.

FAMA, E. (1965): "The Behavior of Stock Market Prices", Journal of Business 38, pp. 34–105.

FIGLEWSKI, S. (1997): "Forecasting Volatility", Financial Markets, Institutions and Instruments, pp. 1–88.

GEYER, A. (1992): "Information, Erwartung, Risiko", VVF: München.

GEYER, A. (1994): "Volatility Estimates of the Vienna stock market", Applied Financial Economics 4, pp. 449–455.

HSIEH, D. (1989): "Testing for Nonlinear Dependence in Daily FX Rates", Journal of Business 62, pp. 339–368.

JORION, P. (1995): "Predicting Volatility in the Foreign Exchange Market", Journal of Finance 50, pp. 507–528.

JORION, P. (1997): Value at Risk, Los Angeles: Irwin Publishing.

KIM, D. and S. KON (1994): "Alternative Models for the Conditional Heteroscedasticity of Stock Returns", Journal of Business 67, pp. 563–598.

LAMOUREUX,C. and W. LASTRAPES (1990): "Persistence in Variance, Structural Change and the GARCH Model", Journal of Business and Economic Statistics, pp. 225–234.

NELSON, D. (1991): "Conditional Heteroscedasticity in Asset Returns: A New Approach", Econometrica 59, pp. 347–70.

PAGAN, A. (1996): "The econometrics of financial markets" Journal of Empirical Finance 3, pp. 15–102.

PAGAN, A. and W. SCHWERT (1990): "Alternative Models for Conditional Stock Volatility", Journal of Econometrics 50, pp. 267–290.

PALM, F. (1996): "GARCH Models of Volatility", in: Maddala, C. S. and C. R. Rao (eds): Handbook of Statistics 14, pp. 209–239.

PEIRO, A. (1994): "The Distribution of Stock Returns: International evidence" Applied Financial Economics 4, pp. 431–439.

SCHMITT, C. (1994): "Volatilitätsprognosen", ZEW working paper no. 7.

TAYLOR, S. (1986): Modeling Financial Time Series, London and Chichester: Wiley.

WEST, K. and D. CHO (1994): "The Predictive Ability of Several Models of Exchange Rate Volatility", NBER Technical Working Paper 152.

WHITE, H. (1980): "A Heteroscedastic-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity", Econometrica 48, pp. 421–448.